

Local average causal effects and superefficiency

Peter M. Aronow*

February 8, 2016

Abstract

Recent approaches in causal inference have proposed estimating average causal effects that are local to some subpopulation, often for reasons of efficiency. These inferential targets are sometimes data-adaptive, in that they are dependent on the empirical distribution of the data. In this short note, we show that if researchers are willing to adapt the inferential target on the basis of efficiency, then extraordinary gains in precision can be obtained. Specifically, when causal effects are heterogeneous, any asymptotically normal and root- n consistent estimator of the population average causal effect is superefficient for a data-adaptive local average causal effect. Our result illustrates the fundamental gain in statistical certainty afforded by indifference about the inferential target.

1 Introduction

When causal effects are heterogeneous, then inferences depend on the population for which causal effects are estimated. Although population average causal effects have traditionally been the inferential targets, recent results have focused on estimating average causal effects that are *local* to some subpopulation for reasons of efficiency. These approaches include trimming observations based on the distribution of the propensity score (Crump et al., 2009), using regression adjustment to estimate reweighted causal effects (Angrist and Pischke, 2009; Humphreys, 2009), or implementing calipers for propensity-score matching (Rosenbaum and Rubin, 1985; Austin, 2011). In some cases, the target parameter is dependent on the empirical distribution of the data, including cases where the researcher is explicitly conducting inference on, e.g., the average treatment effect among the treated conditional on the observed covariate distribution (Abadie and Imbens, 2002), or other causal sample functionals (Aronow, Green, and Lee, 2014; Balzer, Petersen, and van der Laan, 2015), without revision to the estimator being used.

This approach, taken at full generality, implies a form of indifference to which population causal effects are measured for. This indifference can be codified in the form of a *data-adaptive*

*Peter M. Aronow is Assistant Professor, Departments of Political Science and Biostatistics, Yale University, 77 Prospect St., New Haven, CT 06520 (Email: peter.aronow@yale.edu). The author thanks Don Green, Cyrus Samii, Jas Sekhon and Mark van der Laan for helpful comments. All remaining errors are the author's responsibility.

target parameter (van der Laan, Hubbard, and Pajouh, 2015) that is allowed to vary with the data depending on which subpopulation's local average causal effect is best estimated. When treatment effects are heterogeneous, adaptively changing the target parameter on the basis of efficiency yields an unusual result: if the population average causal effect can be consistently estimated with a root- n consistent and asymptotically normal estimator $\hat{\theta}$, then the same estimator $\hat{\theta}$ is always superefficient (i.e., faster than root- n consistent) for a data-adaptive local average causal effect. Furthermore, with an additional regularity condition on mean square convergence, we show that the mean square error of $\hat{\theta}$ for a data-adaptive local average causal effect is of $o(n^{-1})$.

2 Results

Consider a full data probability distribution G with an associated causal effect distribution τ with finite expectation $E_G[\tau]$, where $E_G[\cdot]$ denotes the expectation over the distribution G . We impose a regularity condition on τ establishing non-degeneracy of τ .

Assumption 1 (Effect heterogeneity). $\min(\sup(\text{Supp}(\tau)) - E_G[\tau], E_G[\tau] - \inf(\text{Supp}(\tau))) = c > 0$.

We observe an empirical distribution F_n . Suppose we have an root- n consistent and asymptotically normal estimator of the average causal effect $E_G[\tau]$, $\hat{\theta}$.

Definition 1. An estimator $\hat{\theta}$ is root- n consistent and asymptotically normal for θ_0 if $\sqrt{n}(\hat{\theta} - \theta_0) = \mathcal{N}(0, \sigma^2) + o_p(1)$, for some $0 < \sigma^2 < \infty$.

We now define the target parameter, θ_{F_n} .

Definition 2. Let the target parameter

$$\theta_{F_n} = \begin{cases} \hat{\theta} & : |\hat{\theta} - E_G[\tau]| \leq c \\ E_G[\tau] + c & : \hat{\theta} - E_G[\tau] > c \\ E_G[\tau] - c & : \hat{\theta} - E_G[\tau] < -c \end{cases},$$

where, as in Assumption 1, $c = \min(\sup(\text{Supp}(\tau)) - E_G[\tau], E_G[\tau] - \inf(\text{Supp}(\tau)))$.

The target parameter adapts naturally to the closest value in an interval surrounding $E_G[\tau]$, where the width of the interval is defined by the support of τ . We formalize how each θ_{F_n} is a local average treatment effect.

Proposition 1. There exists a nonnegative weighting associated with each empirical distribution F_n , w_{F_n} , such that across all F_n , $\theta_{F_n} = \frac{E_G[w_{F_n}\tau]}{E_G[w_{F_n}]}$.

A proof of Proposition 1 follows directly from the fact that a weighted mean can obtain any value in the interval defined by the infimum and supremum of its distribution's support. We now prove the the superefficiency of $\hat{\theta}$.

Proposition 2. *Suppose that Assumption 1 holds. Then for any root- n consistent and asymptotically normal estimator of $E_G[\tau]$, $\hat{\theta}$, $\sqrt{n}(\hat{\theta} - \theta_{F_n}) = o_p(1)$.*

Proof. The author thanks Jas Sekhon for suggesting the following proof strategy. Decompose $\hat{\theta}$ into $\tilde{\theta} = \mathcal{N}(E_G[\tau], \sigma^2/n)$ and $u = o_p(n^{-1/2})$, so that $\hat{\theta} = \tilde{\theta} + u$. Since $(\tilde{\theta} - \theta_{F_n})$ is $o_p(a_n)$ for any positive sequence (a_n) , the rate of convergence of $\hat{\theta}$ is at worst governed by the bound ensured by u 's $o_p(n^{-1/2})$ convergence. To prove the claim, note that for any positive ε , $\Pr\left(\frac{|\tilde{\theta} - \theta_{F_n}|}{a_n} \geq \varepsilon\right) \leq \Pr(\tilde{\theta} - \theta_{F_n} \neq 0) = 2\Phi(-c\sqrt{n}/\sigma)$. Since $\lim_{n \rightarrow \infty} 2\Phi(-c\sqrt{n}/\sigma) = 0$, $(\tilde{\theta} - \theta_{F_n})$ is $o_p(a_n)$. Thus $\hat{\theta} - \theta_{F_n} = o_p(a_n) + o_p(n^{-1/2}) = o_p(n^{-1/2})$, yielding the result. \square

When an additional regularity condition is imposed on the convergence of $\hat{\theta}$ to normality, a stronger result can be obtained about the rate of mean square convergence.

Proposition 3. *Suppose that $\hat{\theta}$ obeys $\sqrt{n}(\hat{\theta} - E_G[\tau]) = \mathcal{N}(0, \sigma^2) + \varepsilon$, where $E_G[\varepsilon^2] = o(n^{-1/2})$. Then $E_G[(\hat{\theta} - \theta_{F_n})^2] = o(n^{-1})$.*

Proof. We will show that the mean square error of $(\tilde{\theta} - \theta_{F_n})$ converges to zero sufficiently quickly, implying that the rate of convergence of $\hat{\theta}$ is at worst governed by the mean square error bound ensured by ε 's convergence rate. To obtain the rate of convergence of the mean square error of $\tilde{\theta}$, we integrate over its squared deviation from the target parameter. Within c of $E_G[\tau]$, the squared deviation is zero, thus we need only integrate over the squared deviation over the tails of the normal distribution. To ease calculations, we obtain an upper bound by integrating over the squared deviation from $E_G[\tau]$, rather than from θ_{F_n} :

$$\begin{aligned} E_G[(\tilde{\theta} - \theta_{F_n})^2] &\leq 2 \int_c^\infty x^2 \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{x^2 n}{2\sigma^2}} \\ &= c\sigma \sqrt{\frac{2}{\pi}} \frac{e^{-\frac{c^2 n}{2\sigma^2}}}{\sqrt{n}} + 2\sigma^2 \frac{\Phi\left(\frac{-c\sqrt{n}}{\sigma}\right)}{n} \\ &= o(n^{-1}). \end{aligned}$$

Since $E_G[(\tilde{\theta} - \theta_{F_n})^2] = o(n^{-1})$ and $n^{-1/2}E_G[\varepsilon^2] = o(n^{-1})$, the Cauchy-Schwarz inequality ensures that $E_G[(\hat{\theta} - \theta_{F_n})^2] = o(n^{-1}) + o(n^{-1}) = o(n^{-1})$. \square

3 Discussion

Our results highlight the additional certainty obtained by indifference about the population for which average causal effects are measured. It is well known that efficiency gains may be obtained through data-adaptive inference. But the extent to which the researcher benefits from indifference about the target parameter has been understated. Under treatment effect heterogeneity – a precondition for locality to be a concern – all root- n consistent and asymptotically normal estimators of the average treatment effect are superefficient for a local average treatment effect. And while we do not speak to the substantive implications of indifference in scientific inquiry, we show how such indifference yields greatly increased statistical certainty.

References

- Abadie, A. and Imbens, G. 2002. Simple and bias-corrected matching estimators for average treatment effects. NBER technical working paper no. 283.
- Angrist, J.D. and Pischke, J.S. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.
- Aronow, P.M., Green, D.P. and Lee, D.K.K. 2014. Sharp bounds on the variance in randomized experiments. *Annals of Statistics*. 42(3) 850–871.
- Austin, P.C., 2011. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, 10(2), pp.150–161.
- Balzer, L.B., Petersen, M.L. and van der Laan, M.J. 2015. Targeted estimation and inference for the sample average treatment effect. bepress.
- Crump, R.K., Hotz, V.J., Imbens, G.W. and Mitnik, O.A. 2009. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*.
- Humphreys, M., 2009. Bounds on least squares estimates of causal effects in the presence of heterogeneous assignment probabilities. Manuscript, Columbia University.
- Rosenbaum P.R. and Rubin D.B. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician* 39(1):33–38
- van der Laan, M.J., Hubbard, A.E. and Pajouh, S.K. (2013). Statistical inference for data adaptive target parameters. bepress.